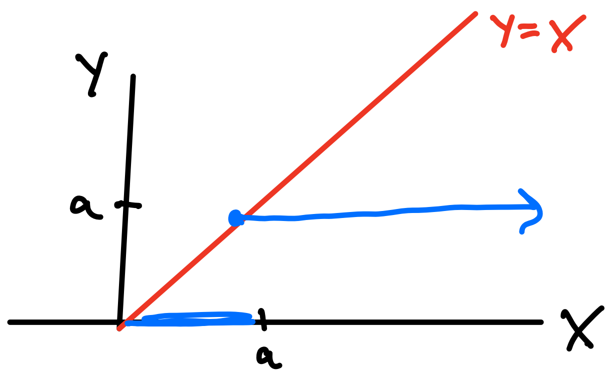
Stat 134 lec 11

Warmup:

This question asks you to graph

Let $Y = \begin{cases} a & \text{if } X \ge a \\ 0 & \text{else} \end{cases}$



(diagonal line drawn for convenience)

Assume $X \ge 0$, $a > 0$

$X \ge \begin{cases} a & \text{if } X \ge a \\ 0 & \text{else} \end{cases} = a \cdot I(X \ge a)$

indicator

$E(X) \ge 0 \cdot P(X < a) + a \cdot P(X \ge a)$

$\boxed{P(X \ge a) \le \frac{E(X)}{a}}$ ← Markov's Inequality

Announcement: Q2 in section next Thursday

Coverage: Sections 2.1, 2.2, 2.4, 2.5, 3.1, 3.2

Stat 134

Chapter 3 **Friday February 15 2019**

1. Consider a well shuffled deck of cards. The expected number of cards before the first ace is?

   a 52/5
   b 48/5
   c 48/4
   d none of the above

d	The p of getting an ace is 4/52 or 1/13, therefore, the EV should be 52/13 or np.

b	There are 48 possible cards to pull before the first Ace (52 cards - 4 Aces), which makes n = 48. For each individual card, there are 5 places it could go: before all the Aces, after the 1st Ace, after the 2nd Ace, after the 3rd Ace, and after the 4th Ace, making p = 1/5. E(X) = np = 48/5

$E(X) = P(X \geq 1) + P(X \geq 2) + P(X \geq 3) + \dots$ Tail Sum Formula

This is useful when X = min or max,

Discrete Distributions

1. Ber (p)
2. Bin (n, p)
3. HG (n, N, G)
4. Pois (μ)
5. Unif $\{1, \dots, n\}$
6. Geom(p) on $\{1, 2, \dots\}$

Geometric RV

\# trials until first success

ex X = number of p coin tosses until your first heads

$X=1$	H	p
$X=2$	TH	qp
$X=3$	TTH	$q^2 p$

$$P(X=k) = q^{k-1} p$$ Geom(p) formula on $\{1, 2, \dots\}$

Note trials are independent

Today

1. Sec 3.2 Markov inequality
2. Sec 3.3 $SD(X)$, $Var(X)$, Chebyshev's Inequality

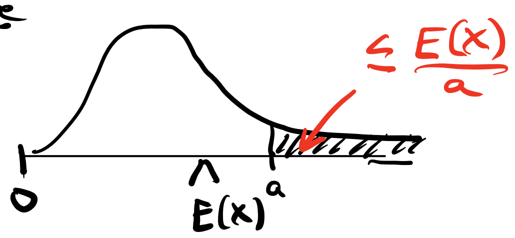
① Sec 3.2 Markov Inequality

Proved in warmup

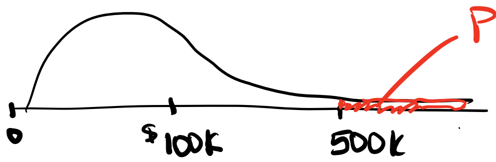
Markov's inequality:

If $X \geq 0$, then $P(X \geq a) \leq \frac{E(X)}{a}$ for every $a > 0$.

Picture



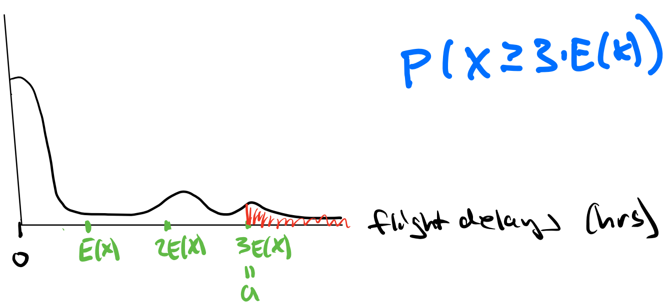
ex Let X be the yearly income of Bay area residents. $E(X) = \$100K$. Find an upper bound for $P(X \geq 500K)$



$P(X \geq 500K) \leq \frac{E(X)}{500K} = \boxed{1/5}$

ex Give an upper bound for the fraction of all US flights that have delay times greater than 3 or more times the national average.

$P(X \geq 3 \cdot E(X)) \leq \frac{E(X)}{3E(X)} = \boxed{1/3}$



ex Let $X_1, X_2, \dots, X_{100}$ be independent and identically distributed (iid) Pois(.01).

Let $S = X_1 + X_2 + \cdots + X_{100}$

$X \sim \text{Pois}(\mu)$

$$P(X=k) = \frac{e^{-\mu}\mu^k}{k!}$$

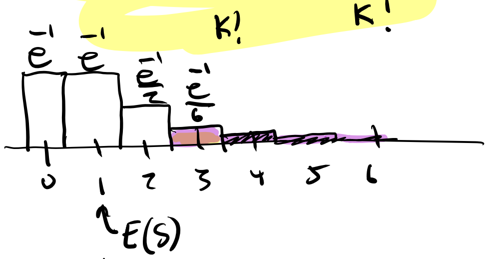
a) What distribution is S? $\text{Pois}(100(.01)) = \text{Pois}(1)$

b) Find an upperbound for $P(S \ge 3)$ using Markov's inequality.

$$P(S \ge 3) \le \frac{\overbrace{E(S)}^{=1}}{3} = \boxed{1/3}$$

Note Exact: $P(S \ge 3) = 1 - \underset{e^{-1}}{\underset{\parallel}{P(0)}} - \underset{e^{-1}}{\underset{\parallel}{P(1)}} - \underset{\frac{e^{-1}}{2}}{\underset{\parallel}{P(2)}}$

$$P(S=k) = \frac{e^{-1}1^k}{k!} = \frac{e^{-1}}{k!}$$



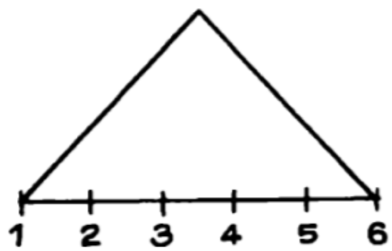
$= 1 - e^{-1}(1+1+\frac{1}{2})$

$= .08$

② <u>Sec 3.3</u> <u>Standard deviation</u> (SD)

SD is the average spread of your data around the mean.

What is the SD of the following figure?



a 0.5

b 1

c 2

$$SD(x) = \sqrt{E\left((x - E(x))^2\right)}$$

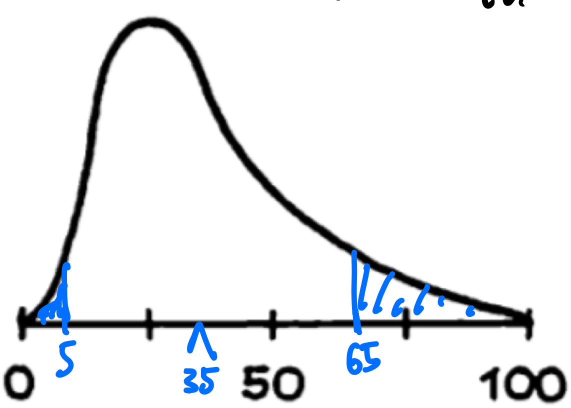
$$Var(x) = (SD(x))^2 = E\left((x - E(x))^2\right)$$

Chebyshev's Inequality

For any random variable X, and any $K>0$,

$$P\left(|X - E(X)| \geq K \cdot SD(X) \right) \leq \frac{1}{K^2}$$

ex Let X have distribution with $E(X) = 35$, $SD(X) = 15$.



$2 \cdot 15 \Rightarrow \boxed{K=2}$

Find $P(|X-35| \geq 30)$?

$$\leq \frac{1}{2^2} = \frac{1}{4}$$

What can you say about $P(X \geq 65)$? $\leq 1/4$

M: $P(X \geq 65) \leq \frac{35}{65}$

less info.

tinyurl.com/feb10-2023

Stat 134

1. A list of non negative numbers has an average of 1 and an SD of 2. Let p be the proportion of numbers greater than or equal to 5. To get an upper bound for p, you should:

   ~~**a**~~ Assume a normal distribution

   b Use Markov's inequality $a=5$ M: $P(X \geq 5) \leq \frac{1}{5}$

   c Use Chebyshev's inequality C: $P(X \geq \underset{1+2\cdot 2}{5}) \leq \frac{1}{4}$

   d none of the above

$$P(|X-1| \geq \overset{2\cdot 2}{4}) = \underbrace{P(X \geq 5)}_{1+2\cdot 2} + \underbrace{P(X \leq -3)}_{=0}$$

Proof of Chebyshev

For any random variable X, and any K>0

$$P\left(|X-E(X)| \geq K\,SD(X)\right) \leq \frac{1}{K^2}$$

By Markov

$$P(Y \geq a) \leq \frac{E(Y)}{a} \quad \text{for } \begin{array}{l} Y \geq 0 \\ a > 0 \end{array}$$

Note that $P(A^2 < B^2) = P(|A| < |B|)$

since $A^2 < B^2$ iff $|A| < |B|$ for numbers A, B.

Hence,

$$P\left(|X-E(X)| \geq K\,SD(X)\right)$$

(Here $A = X-E(X)$, $B = K\,SD(X)$)

$$= P\left((X-E(X))^2 \geq (K\,SD(X))^2\right) \leq \frac{E\left((X-E(X))^2\right)}{(K\,SD(X))^2}$$

$$= \frac{\cancel{Var(X)}}{K^2\,\cancel{Var(X)}} = \frac{1}{K^2}$$

$\square$